Homogenization of brain MRI from a clinical data warehouse using contrast-enhanced to non-contrast-enhanced image translation with U-Net derived models

Simona Bottani, Elina Thibeau-Sutre, Aurélien Maire, Sebastian Ströer, Didier Dormont, Olivier Colliot, Ninon Burgos

Homogenization of brain MRI from a clinical data warehouse using contrast-enhanced to non-contrast-enhanced image translation with U-Net derived models

Simona Bottani[a], Elina Thibeau-Sutre[a], Aurélien Maire[b], Sebastian Ströer[c], Didier Dormont[a,c], Olivier Colliot[a], Ninon Burgos[a], and the APPRIMAGE Study Group[1]

[a]Inria, Aramis project-team, Sorbonne Université, Institut du Cerveau - Paris Brain Institute, Inserm, CNRS, AP-HP, Hôpital Pitié Salpêtrière, Paris, France
[b]AP-HP, WIND department, Paris, France
[c]AP-HP, Hôpital Pitié Salpêtrière, Department of Neuroradiology, Paris, France

ABSTRACT

Clinical data warehouses provide access to massive amounts of medical images and thus offer unprecedented opportunities for research. However, they also pose important challenges, a major challenge being their heterogeneity. In particular, they contain patients with numerous different diseases. The exploration of some neurological diseases with magnetic resonance imaging (MRI) requires injecting a gadolinium-based contrast agent (for instance to detect tumors or other contrast-enhancing lesions) while other diseases do not require such injection. Image harmonization is a key factor to enable unbiased differential diagnosis in such context. Additionally, classical neuroimaging software tools that extract features used as inputs of classification algorithms are typically applied only to images without gadolinium. The objective of this work is to homogenize images from a clinical data warehouse and enable the extraction of consistent features from brain MR images, no matter the initial presence or absence of gadolinium. We propose a deep learning approach based on a 3D U-Net to translate contrast-enhanced into non-contrast-enhanced T1-weighted brain MRI. The approach was trained/validated using 230 image pairs and tested on 26 image pairs of good quality and 51 image pairs of low quality from the data warehouse of the hospitals of the Greater Paris area (Assistance Publique-Hôpitaux de Paris [AP-HP]). We tested two different 3D U-Net architectures and we chose the one reaching the best image similarity metrics for a further validation for a segmentation task. We tested two 3D U-Net architectures with the addition either of residual connections or of attention mechanisms. The U-Net with attention mechanisms reached the best image similarity metrics and was further validated on a segmentation task. We showed that features extracted from the synthetic images (gray matter, white matter and cerebrospinal fluid volumes) were closer to those obtained from the non-contrast-enhanced T1-weighted brain MRI (considered as reference) than the original, contrast-enhanced, images.

1. INTRODUCTION

Clinical data warehouses, gathering hundreds of thousands of medical images from numerous hospitals, offer fantastic opportunities for computer-aided diagnosis for neurological diseases but also pose considerable challenges. One of these challenges is the heterogeneity of the images present in the data warehouse. Neurological diseases can result in a variety of brain lesions that are each studied with specific magnetic resonance imaging (MRI) sequences. For example, T1-weighted (T1w) brain MRI enhanced with a gadolinium-based contrast agent are used for the study of lesions such as tumors, and T1w images without gadolinium are used for the study of neurodegenerative diseases. To perform differential diagnosis using classification algorithms, homogeneous features must be extracted from the images, no matter the disease, otherwise a link could be established between MRI sequence and pathology, which would create bias. Software tools such as SPM,[1] ANTs[2] or FSL[3] have been widely used for feature extraction but they were largely validated using structural T1w MRI without gadolinium only,

Further author information: (Send correspondence to Simona Bottani)
A.A.A.: E-mail: simonabottani92@gmail.com

to the best of our knowledge, and their good performance on images with gadolinium is thus not guaranteed. A solution could then be to convert contrast-enhanced T1w (T1w-ce) into non-contrast-enhanced T1w (T1w-nce) brain MRI before using such tools.

Deep learning has been widely used in the image translation domain to enhance image quality, such as for denoising[4–7] or super-resolution,[8–12] but also for image harmonization.[13] Closer to our objective, 3D U-Net like models have been developed for the synthesis of images with gadolinium from images without gadolinium.[14–16]

Our objective in this work is to obtain a homogeneous dataset of T1w-nce images that will be later used for the differential diagnosis of neurological diseases in a clinical data warehouse. We propose two deep learning models, based on a 3D U-Net like structure that demonstrated good performance on similar tasks,[14–16] to translate T1w-ce into T1w-nce images. We trained and tested our models using 307 pairs of T1w-nce and T1w-ce images. We first assessed synthesis accuracy by comparing real and synthetic T1w-nce images using standard metrics. We then compared the volumes of different tissues obtained by segmenting the real T1w-nce, real T1w-ce and synthetic T1w-nce images using SPM.[17]

2. MATERIALS AND METHODS

2.1 Dataset description

This work relies on a large clinical dataset containing all the T1w brain MR images of adult patients scanned in hospitals of the Greater Paris areas (Assistance Publique-Hôpitaux de Paris [AP-HP]). The data were made available by the AP-HP data warehouse and the study was approved by the Ethical and Scientific Board of the AP-HP. According to French regulation, consent was waived as these images were acquired as part of the routine clinical care of the patients. Images from this clinical data warreehouse are very heterogeneous:[18] they include images of patients with a wide range of ages and diseases, acquired with different scanners from 1980 up to now.

The dataset used in this work is composed of 307 pairs of T1w-ce and T1w-nce images that were extracted from a dataset composed of 9941 images made available by the AP-HP data warehouse. We first selected all the images of low, medium and good quality, excluding images that were not proper T1-weighted brain MRI,[18] resulting in 7397 images. This selection was based on manual quality control for 5500 images and on automatic quality control for the remaining 4441 images.[18] We then considered only patients having one T1w-ce and one T1w-nce at the same session, with a T1w-nce image of medium or good quality. Finally, we visually checked all the images and excluded 52 image pairs that were potential outliers because of extremely large lesions. Among the selected images we had 256 image pairs of medium and good quality, and 51 image pairs with a T1w-ce of low quality and a T1w-nce of good or medium quality.

2.2 Image preprocessing

All the images were organised using the Brain Imaging Data Structure (BIDS).[19] We applied the following pre-processing using the 't1-linear' pipeline of Clinica,[20] which is a wrapper of the ANTs software.[2] Bias field correction was applied using the N4ITK method.[21] An affine registration to MNI space was performed using the SyN algorithm.[22] The registered images were further rescaled based on the min and max intensity values, and cropped to remove background resulting in images of size $169\times208\times179$, with 1 mm isotropic voxels.[23] Finally all the images were resampled to have a size of $128\times128\times128$ using trilinear interpolation.

2.3 Network architecture

We implemented two 3D U-Net like models: *Res-U-Net* with residual connections[14,24] and *Att-U-Net* with attention mechanisms[25] (Fig. 1). The U-Net structure allows preserving the original structure of the images thanks to the skip connections while the residual connections and the attention gates are both useful to make the model focus on the salient parts of the images. We used the ADAM optimizer, the L1 loss, a batch size of 2, and we trained the models for 300 epochs. We relied on Pytorch for the implementation.

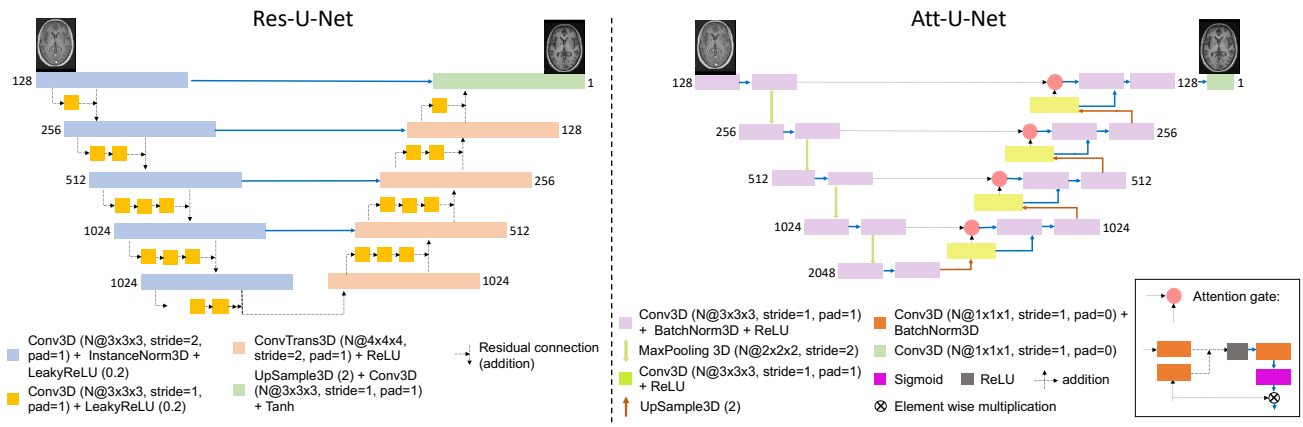

Figure 1. Architectures of the proposed 3D U-Net models. The models take as input the real T1w-nce image of size 128×128×128 and generate the synthetic T1w-nce of size 128×128×128. *Res-U-Net* (left): images pass through five descending blocks, each one followed by a residual module, and then through four ascending blocks and one final layer. *Att-U-Net* (right): images pass through five descending blocks and then through four ascending blocks and one final layer. One of the input of each ascending block is the results of the attention gate composed of convolutional blocks, ReLU and sigmoid layers. All the parameters such as kernel size, stride, padding, size of each feature map (N) are reported.

2.4 Experiments and validation of the model

The experiments relied on 307 pairs of T1w-ce and T1w-nce images. We randomly selected 10% of the 256 image pairs of medium and good quality for testing (dataset called $\text{Test}_{\text{good}}$), the other 230 image pairs being used for training/validation. Only images of good and medium quality were used for training/validation to ensure that the model focuses on the differences related to the presence or absence of gadolinium, and not to other factors. The remaining 51 image pairs with a T1w-ce of low quality and a T1w-nce of good or medium quality were used only for testing (dataset called Test_{low}).

2.4.1 Synthesis accuracy

Image similarity was evaluated using the mean absolute error (MAE), peak signal-to-noise ratio (PNSR) and structural similarity (SSIM).[26] We calculated these metrics both between the real and synthetic T1w-nce images and between the real T1w-nce and T1w-ce (as reference). These metrics were calculated within the brain region obtained by skull-stripping the T1w-nce and T1w-ce[27] and computing the union of the two resulting brain masks.

2.4.2 Segmentation fidelity

Our goal is to obtain gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) segmentations from T1w-ce images using widely-used software tools that are consistent with segmentations obtained from T1w-nce images. We thus assessed segmentation consistency by analyzing the tissue volumes resulting from the segmentations, which are important features when studying atrophy in the context of neurodegenerative diseases.

We processed the images using the 't1-volume-tissue-segmentation' pipeline of Clinica.[20,28] This wrapper of the Unified Segmentation procedure implemented in SPM[17] simultaneously performs tissue segmentation, bias correction and spatial normalization. The probability maps were then binarized to derive tissue volumes. We calculated the volume difference for each tissue between the T1w-nce and the synthetic T1w-nce or T1w-ce. We multiplied the difference with the average total intracranial volume computed across the two test sets. To assess whether the tissue volumes presented a statistically significant difference depending on the images they were obtained from, we performed paired t-tests correcting for multiple comparisons using the Bonferroni method.

3. RESULTS

We report results of the proposed 3D U-Net like models trained on 230 image pairs and tested on $\text{Test}_{\text{good}}$ and Test_{low} obtained from a clinical multisite dataset (examples in Fig 2). We note the absence of contrast agent in the synthetic T1w-nce, while it is clearly visible in the sagittal slice of the T1w-ce, and that the anatomical

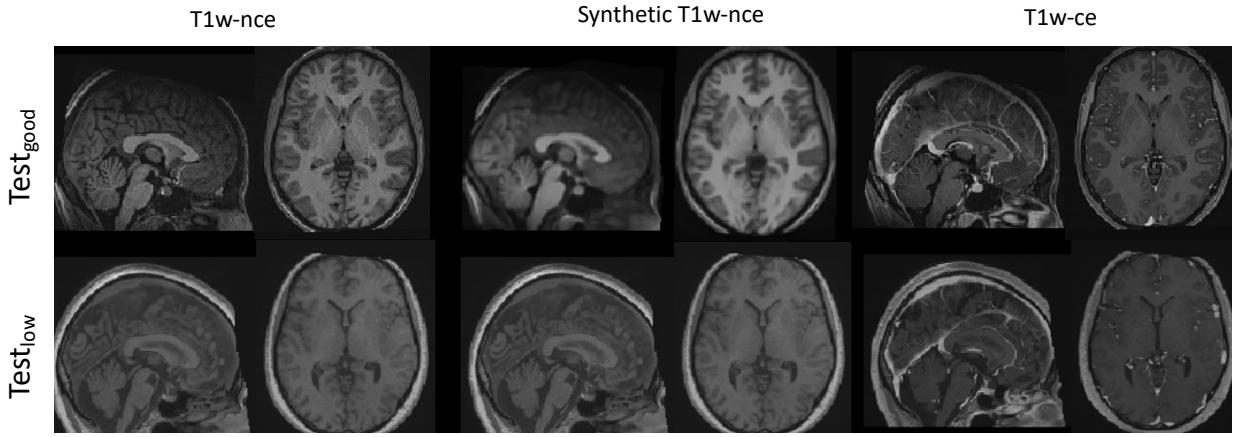

Figure 2. Examples of real T1w-nce, synthetic T1w-nce (obtained with the *Att-U-Net* model) and real T1w-ce images for two patients (at the top from $\text{Test}_{\text{good}}$ and at the bottom from Test_{low}) in the sagittal and axial planes.

Table 1. MAE, PSNR and SSIM obtained on the two independent test sets with various image quality. For each metric, we report the average and standard deviation across the corresponding test set. We compute the metrics for both T1w-ce and synthetic T1w-nce in relation to the real T1w-nce, and so within the brain region.

Test set	Compared images	Model	MAE (%)	PSNR (dB)	SSIM
$\text{Test}_{\text{good}}$	T1w-nce / T1w-ce	-	4.14 ± 1.59	23.03 ± 2.83	0.90 ± 0.05
	T1w-nce / Synthetic T1w-nce	*Att-U-Net*	2.73 ± 1.69	29.07 ± 4.53	0.96 ± 0.05
		Res-U-Net	3.06 ± 1.50	26.89 ± 4.30	0.95 ± 0.04
Test_{low}	T1w-nce / T1w-ce	-	3.71 ± 1.99	24.20 ± 3.85	0.91 ± 0.06
	T1w-nce / Synthetic T1w-nce	*Att-U-Net*	2.89 ± 1.85	27.15 ± 4.57	0.95 ± 0.05
		Res-U-Net	2.93 ± 1.77	26.71 ± 4.32	0.95 ± 0.05

structures are preserved between the synthetic and real T1w-nce. We also note that for the patient from Test_{low} the contrast between gray and white matter seems improved in the synthetic compared with the real T1w-ce.

3.1 Synthesis accuracy

Table 1 reports the image similarity metrics obtained for the two test sets within the brain region. We computed these metrics to assess the similarity between real and synthetic T1w-nce images, but also between T1w-nce and T1w-ce images to set a baseline. We observe that for both models, the similarity is higher between real and synthetic T1w-nce images than between T1w-nce and T1w-ce images according to all three metrics on both test sets. *Att-U-Net* performed slightly but systematically better than *Res-U-Net* and was thus kept in the following.

3.2 Segmentation fidelity

Absolute volume differences obtained between T1w-nce and T1w-ce images and between T1w-nce and synthetic T1w-nce images are reported in Table 2. For both test sets and all tissues, the absolute volume differences are smaller between real and synthetic T1w-nce images than between T1w-nce and T1w-ce images. The differences between T1-nce/synthetic T1w-nce and T1w-nce/T1w-ce are statistically significant (corrected p-value <10e-3) for GM and CSF when considering $\text{Test}_{\text{good}}$, and for all tissues when considering Test_{low}.

4. DISCUSSION AND CONCLUSION

The use of clinical images for the validation of computer-aided diagnosis (CAD) systems is still largely unexplored. One of the obstacles lies in the heterogeneity of the data acquired in the context of routine clinical practice. Homogenization of the dataset, and consequently of the features extracted from it, is an important step for the development of reliable CAD systems in a clinical setting. To homogenize such dataset, we proposed U-Net like structures to synthesize T1w-nce from T1w-ce images and we trained the models using images from a clinical data warehouse. To the best of our knowledge, this is the first work to propose such approach in this context.

Table 2. Absolute volume difference (mean ± standard deviation in cm^3) between T1w-nce and T1w-ce images and between T1w-nce and synthetic T1w-nce images (obtained with the *Att-U-Net* model) for GM, WM and CSF. The corresponding corrected p-values were obtained with paired t-tests corrected for multiple comparisons using the Bonferroni correction.

Test set	Compared images	Gray matter		White matter		CSF	
		Absolute difference [cm^3]	P-value	Absolute difference [cm^3]	P-value	Absolute difference [cm^3]	P-value
$Test_{good}$	T1w-nce / T1w-ce	26.68 ± 15.92	<10e-3	10.81 ± 3.71	0.12	61.62 ± 34.61	<10e-3
	T1w-nce / Synthetic T1w-nce	10.36 ± 6.98		7.79 ± 5.87		13.37 ± 10.18	
$Test_{low}$	T1w-nce / T1w-ce	49.63 ± 49.38	<10e-3	25.36 ± 27.73	<10e-3	69.55 ± 37.77	<10e-3
	T1w-nce / Synthetic T1w-nce	19.61 ± 29.54		13.95 ± 24.74		18.27 ± 17.20	

We first showed using MAE, PSNR and SSIM that the similarity between real and synthetic T1w-nce images was higher than the similarity between real T1w-nce and T1w-ce images, no matter the U-Net model. The synthesis accuracy was of the same order as that of recent works,[14,15] but slightly better performance was reached with the *Att-U-Net* model, which was thus further evaluated. In the second step of the validation, we showed that the absolute volume differences of GM, WM and CSF were larger between real T1w-nce and Tw-ce images than between real and synthetic T1w-nce images (statistically significant difference most of the times). This confirms the hypothesis that gadolinium-based contrast agent may alter the contrast between the different brain tissues, making features extracted from such images with standard segmentation tools, here SPM,[1] unreliable. At the same time, we validated the suitability of the synthetic images since their segmentation was consistent with those obtained from real T1w-nce images as the volume differences were small. Finally, our results show that the proposed model performs well, no matter the quality of the input T1w-ce image, even though more accurate results are reached for images of good/medium quality.

Several steps remain to be performed before using synthetic T1-nce images for the differential diagnosis of neurological diseases. For example, the performance of CAD systems trained with a mix of real T1w-nce and T1w-ce images should be compared with that of CAD systems trained with a mix of real and synthetic T1w-nce images. Future work could also involve implementing and evaluating other approaches, such as generative adversarial networks, to further improve the quality of the synthetic images, in particular their sharpness.

ACKNOWLEDGMENTS

The research was done using the Clinical Data Warehouse of the Greater Paris University Hospitals. The authors are grateful to the members of the AP-HP WIND and URC teams, and in particular Stéphane Bréant, Florence Tubach, Jacques Ropers, Antoine Rozès, Camille Nevoret, Christel Daniel, Martin Hilka, Yannick Jacob, Julien Dubiel and Cyrina Saussol. They would also like to thank the "Collégiale de Radiologie of AP-HP" as well as, more generally, all the radiology departments from AP-HP hospitals.
The research leading to these results has received funding from the Abeona Foundation (project Brain@Scale), from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6).

REFERENCES

[1] Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E., [*Statistical parametric mapping: the analysis of functional brain images*], Elsevier (2011).

[2] Avants, B. B., Tustison, N. J., Stauffer, M., Song, G., Wu, B., and Gee, J. C., "The Insight ToolKit image registration framework," *Frontiers in Neuroinformatics* **8**, 44 (2014).

[3] Mark, J., Christian, F. B., Timothy, E. B., Mark, W. W., and Stephen, M. S., "FSL," *NeuroImage* **62**(2), 782–790 (2012).

[4] Hashimoto, F., Ohba, H., Ote, K., Teramoto, A., and Tsukada, H., "Dynamic PET image denoising using deep convolutional neural networks without prior training datasets," *IEEE Access* **7**, 96594–96603 (2019).

[5] Benou, A., Veksler, R., Friedman, A., and Raviv, T. R., "Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences," *Medical Image Analysis* **42**, 145–159 (2017).

[6] Jiang, D., Dou, W., Vosters, L., Xu, X., Sun, Y., and Tan, T., "Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network," *Japanese journal of radiology* **36**(9), 566–574 (2018).

[7] Yang, Z., Zhuang, X., Sreenivasan, K., Mishra, V., Curran, T., and Cordes, D., "A robust deep neural network for denoising task-based fMRI data: An application to working memory and episodic memory," *Medical Image Analysis* **60**, 101622 (2020).

[8] Chen, Y., Shi, F., Christodoulou, A. G., Xie, Y., Zhou, Z., and Li, D., "Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 91–99, Springer (2018).

[9] Du, J., Wang, L., Liu, Y., Zhou, Z., He, Z., and Jia, Y., "Brain mri super-resolution using 3d dilated convolutional encoder–decoder network," *IEEE Access* **8**, 18938–18950 (2020).

[10] Kim, K. H., Do, W.-J., and Park, S.-H., "Improving resolution of MR images with an adversarial network incorporating images with different contrast," *Medical Physics* **45**(7), 3120–3131 (2018).

[11] Pham, C.-H., Ducournau, A., Fablet, R., and Rousseau, F., "Brain MRI super-resolution using deep 3D convolutional networks," in [*2017 IEEE ISBI*], 197–200 (2017).

[12] Zeng, K., Zheng, H., Cai, C., Yang, Y., Zhang, K., and Chen, Z., "Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network," *Computers in Biology and Medicine* **99**, 133–141 (2018).

[13] Dewey, B. E., Zhao, C., Reinhold, J. C., Carass, A., Fitzgerald, K. C., Sotirchos, E. S., Saidha, S., Oh, J., Pham, D. L., Calabresi, P. A., van Zijl, P. C. M., and Prince, J. L., "DeepHarmony: a deep learning approach to contrast harmonization across scanner changes," *Magnetic Resonance Imaging* **64**, 160–170 (2019).

[14] Bône, A., Ammari, S., Lamarque, J.-P., Elhaik, M., Chouzenoux, É., Nicolas, F., Robert, P., Balleyguier, C., Lassau, N., and Rohé, M.-M., "Contrast-enhanced brain MRI synthesis with deep learning: key input modalities and asymptotic performance," in [*2021 IEEE ISBI*], (2021).

[15] Kleesiek, J., Morshuis, J. N., Isensee, F., Deike-Hofmann, K., Paech, D., Kickingereder, P., Köthe, U., Rother, C., Forsting, M., Wick, W., Bendszus, M., Schlemmer, H.-P., and Radbruch, A., "Can virtual contrast enhancement in brain MRI replace gadolinium?: a feasibility study," *Investigative Radiology* **54**(10), 653–660 (2019).

[16] Sun, H., Liu, X., Feng, X., Liu, C., Zhu, N., Gjerswold-Selleck, S. J., Wei, H.-J., Upadhyayula, P. S., Mela, A., Wu, C.-C., Canoll, P. D., Laine, A. F., Vaughan, J. T., Small, S. A., and Guo, J., "Substituting Gadolinium in Brain MRI Using DeepContrast," in [*2020 IEEE ISBI*], 908–912 (2020).

[17] Ashburner, J. and Friston, K. J., "Unified segmentation," *NeuroImage* **26**(3), 839–851 (2005).

[18] Bottani, S., Burgos, N., Maire, A., Wild, A., Ströer, S., Dormont, D., and Colliot, O., "Automatic quality control of brain T1-weighted magnetic resonance images for a clinical data warehouse," *arXiv:2104.08131* (2021).

[19] Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., Poline, J.-B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J. A., Varoquaux, G., and Poldrack, R. A., "The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments," *Scientific data* **3**(1), 1–9 (2016).

[20] Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., Fontanella, S., Gori, P., Guillon, J., Guyot, A., Hassanaly, R., Jacquemont, T., Lu, P., Marcoux, A., Moreau, T., Samper-González, J., Teichmann, M., Thibeau-Sutre, E., Vaillant, G., Wen, J., Wild, A., Habert, M.-O., Durrleman, S., and Colliot, O., "Clinica: An Open-Source Software Platform for Reproducible Clinical Neuroscience Studies," *Frontiers in Neuroinformatics* **15**, 39 (2021).

[21] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C., "N4ITK: improved N3 bias correction," *IEEE Transactions on Medical Imaging* **29**(6), 1310–1320 (2010).

[22] Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C., "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Medical Image Analysis* **12**(1), 26–41 (2008).

[23] Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., and Colliot, O., "Convolutional Neural Networks for Classification of Alzheimer's Disease: Overview and Reproducible Evaluation," *Medical Image Analysis* **63**, 101694 (2020).

[24] Milletari, F., Navab, N., and Ahmadi, S.-A., "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in [*2016 fourth international conference on 3D vision (3DV)*], 565–571, IEEE (2016).

[25] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D., "Attention u-net: Learning where to look for the pancreas," in [*Conference on Medical Imaging with Deep Learning (MIDL 2018)*], (2018).

[26] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).

[27] Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.-P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K. H., and Kickingereder, P., "Automated brain extraction of multisequence MRI using artificial neural networks," *Human Brain Mapping* **40**(17), 4952–4964 (2019).

[28] Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., Routier, A., Guillon, J., Bacci, M., Wen, J., Bertrand, A., Bertin, H., Habert, M.-O., Durrleman, S., Evgeniou, T., and Colliot, O., "Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data," *NeuroImage* **183**, 504–521 (2018).